Acta Cryst. (1998). D**54**, 1453–1455

Bacillus subtilis regulatory protein GerE

Valérie M.-A. Ducros, James A. Brannigan,* Richard J. Lewis and Anthony J. Wilkinson *at Department of Chemistry, University of York, Heslington, York YO10 5DD, England. E-mail: jab@yorvic.york.ac.uk*

(*Received 11 March 1998; accepted 1 April 1998*)

Abstract

GerE is the latest-acting of a series of factors which regulate gene expression in the mother cell during sporulation in *Bacillus*. The gene encoding GerE has been cloned from *B. subtilis* and overexpressed in *Escherichia coli*. Purified GerE has been crystallized by the hanging-drop vapour-diffusion method using polyethylene glycol as a precipitant. The small plate-like crystals belong to the monoclinic space group *C*2 and diffract beyond 2.2 Å resolution with a synchrotron radiation X-ray source.

1. Introduction

Sporulation in *Bacillus subtilis* involves an asymmetric cell division giving rise to two compartments, a mother cell and a smaller forespore, each of which carries a complete copy of the chromosome. Following septation, different programmes of gene expression become activated in the two compartments, orchestrated by regulatory factors which direct the transcription of distinct sets of genes which determine their different developmental fates. The forespore is subsequently engulfed by the mother cell and ultimately matures into a resistant spore which can survive adverse environmental conditions (Errington, 1993; Stragier & Losick, 1996). This resistance is conferred, in part, by the tough protein shell, known as the coat, which surrounds the mature spore.

The biosynthesis of the coat that encases the endospores of Gram-positive bacteria requires the synthesis and assembly of more than a dozen different structural coat proteins (Stragier & Losick, 1996). Transcription of the genes encoding these proteins is temporally controlled by a hierarchical regulatory cascade of four mother cell-specific regulatory factors, two RNA-polymerase sigma factors and two DNA-binding proteins, appearing in the sequence: σ^{E}, SpoIIID, σ^{K} and GerE (Zheng & Losick, 1990; Losick & Stragier, 1992).

The 74-amino-acid DNA-binding protein GerE is the latest-acting of these four regulatory factors. Expression of *gerE* is activated at a late stage of sporulation (stage V) by the appearance of the sigma factor σ^{K}. GerE controls the synthesis and assembly of the protein coat. It acts either as a repressor or an activator of a number of genes, notably the *cot* genes encoding the structural components of the coat. GerE is known to inhibit the transcription of *cotA*, *cotM* and *sigK* (Cutting *et al.*, 1989; Henriques *et al.*, 1997; Sandman *et al.*, 1988; Zheng *et al.*, 1992) and stimulate transcription of *cotD, cotB, cotC, cotG* and *cotX* in conjuction with σ^{K} (Zheng *et al.*, 1992; Zhang *et al.*, 1994; Sacco *et al.*, 1995). Null mutations in *gerE* produce misformed spores whose coats are aberrant in ultrastructure and protein composition. Such spores are sensitive to lysozyme and unable to germinate correctly (Feng & Aronson, 1986; Jenkinson & Lord, 1983; Moir, 1981).

Little is known about the structural basis of DNA recognition by GerE, other than that it has been predicted to contain a helix–turn–helix motif (Holland *et al.*, 1987) and that a consensus DNA sequence for GerE binding has been recognized (Zheng *et al.*, 1992; Zhang *et al.*, 1994). In this paper we report the expression, crystallization and X-ray diffraction analysis of GerE from *B. subtilis* as part of a series of studies by our group to dissect the structural basis for the control of sporulation in this organism.

2. Experimental

2.1. *Cloning and overexpression*

For high-level protein expression, *gerE* was cloned into the T7-promoter-based expression vector pET26b (Novagen) and expressed in *E. coli* BL21(DE3), a strain with an inducible T7 RNA polymerase (Studier & Moffatt, 1986). PCR with high-fidelity DNA polymerase (Pfu, Stratagene) was used to amplify *gerE* from *B. subtilis* IG20 (168*trpC2*) chromosomal DNA. Primers were designed to introduce convenient restriction sites for cloning (*Nde*I and *Bam*HI) and to modify the initiation codon from TTG to ATG. The DNA sequence of the resultant expression clone (Accession AJ223964) revealed codon 11 to be CTG (Leu), confirming that predicted by Cutting *et al.* (1989) and concurring with Leu11 in the homologous sequence from *B. stearothermophilus* (Brannigan, unpublished). This is in conflict with both the original sequence which predicts TCG (Ser) in this position (Cutting & Mandelstam, 1986) and with the current version of the genomic sequence (Kunst *et al.*, 1997).

This construct led to high levels of soluble protein production in suitable *E. coli* expression strains (Fig. 1). Cell cultures were grown in LB media containing 30 μg ml^{-1} kanamycin. At an optical density of 0.8 at 600 nm, protein expression was induced by the addition of IPTG to a final concentration of 1 m*M*. After 4 h further growth, the cells were harvested by centrifugation and the cell pellet frozen at 193 K until use.

2.2. *Protein preparation*

The cells were resuspended in 20 ml of buffer *A* containing 50 m*M* Tris–HCl, pH 7.5, 100 m*M* NaCl, 10 m*M* edta and 1 m*M* dithiothreitol (DTT), and subsequently broken by sonication at 277 K in an MSE ultrasonicator. The lysate was clarified by centrifugation at 18000 r min^{-1} for 60 min. GerE was fractionated using ammonium sulfate with GerE precipitating between 50 and 70% saturation, as indicated by SDS–PAGE analysis of the protein fractions. The GerE-containing fraction was resuspended in 50 ml of buffer *A* and applied to a Heparin–Sepharose column, pre-equilibrated with buffer *A*. It was eluted using a linear gradient from 0.1 to 1 *M* NaCl. The GerE-containing fractions, identified by SDS–PAGE, were pooled and dialysed overnight against 10 m*M* *N*-morpholino-

propanesulfonic acid (MOPS) buffer, pH 5.6, containing 1 m*M* DTT. GerE was subsequently purified by ion-exchange chromatography on Pharmacia Mono-S resin and eluted using a gradient from 100 m*M* to 1 *M* NaCl in 25 m*M* MOPS buffer, pH 5.6. The GerE-containing fractions were further purified by gel filtration (Pharmacia S75 10/30) in 50 m*M* Tris–HCl, pH 8.0, and 250 m*M* NaCl. This procedure resulted in an essentially pure GerE preparation as judged by SDS–PAGE (Fig. 1) and yielded approximately 10 mg of GerE per litre of cells.

2.3. *Crystallization, data collection and processing*

GerE was concentrated to approximately 10 mg ml^{-1} in water. All crystallizations were performed by vapour-phase diffusion using the hanging-drop technique in Falcon 3047 multiwell plates at 291 K. 2 µl hanging drops consisted of 1 µl of protein sample with the addition of 1 µl reservoir buffer. Initial crystallization conditions were determined following a sparse-matrix-screening approach (Jancarik & Kim, 1991). In order to facilitate cryo-crystallographic data collection, crystals were transferred directly into a cryo-protectant solution containing 15% glycerol, 30%(*w*/*v*) PEG 4000, 0.1 *M* sodium acetate, pH 5.0, and 0.2 *M* lithium sulfate. Crystals were subsequently mounted in a rayon fibre loop and transferred into a stream of boiling liquid nitrogen. Native data were collected from a single crystal flash-frozen at 100 K at Daresbury Synchrotron Radiation Source, beamline PX 9.6. A total of 174° of data, to a resolution of 2.15 Å, were collected with an oscillation range of 1.0° per image. Data were integrated with *DENZO* and scaled and reduced with *SCALEPACK* (Otwinowski, 1993). Subsequent calculations were performed with the CCP4 program package (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Crystals of GerE with a plate-like morphology grow from 20–30% PEG 4000, 0.1 *M* sodium acetate buffer, pH 5.0, in the presence of 0.2 *M* lithium sulfate or 0.2 *M* ammonium sulfate, over a period of 5–6 days (Fig. 2). The resulting crystals are both fragile and radiation sensitive. Data collection was therefore performed under cryogenic conditions. GerE crystals belong to the monoclinic space group *C*2 with unit-cell dimensions a = 108.74, b = 61.68, c = 71.65 Å and β = 97.15°. The final data are 99.2% complete to 2.15 Å resolution with an overall $R_{\rm merge}$ ($\sum_{hkl}\sum_i |I_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_i \langle I_{hkl}\rangle$) of 0.068, a mean $I/\sigma(I)$ of 11.53 and a multiplicity of 3.2 observations per reflection. In the outer resolution shell (2.32–2.15 Å) $R_{\rm merge}$ is 0.276 with an $I/\sigma(I)$ of 4.07, a completeness of 95.0% and a mean multiplicity of 2.95.

The GerE monomer has a calculated molecular mass of 8582 Da, and gel-filtration analysis indicates that the protein exists as a dimer in solution (data not shown). It is difficult to state categorically the number of molecules in the asymmetric unit, since packing densities for 3–7 monomers in the asymmetric unit all lie within the range observed for protein crystals (Matthews, 1968). These give rise to densities from 4.7–2.0 Å^3 Da^{-1}, corresponding to 73–38% solvent. Self-rotation functions calculated with a variety of input parameters failed to indicate unambiguously the presence of non-crystallographic symmetry elements. The native Patterson function, calculated at 6 Å resolution, revealed a peak with height approximately 10% of the origin peak height at position u = 0.09, v = 0.00 and w = 0.52, which may indicate that at least two of the molecules within the asymmetric unit lie in approximately the same orientation.

The GerE sequence contains a putative helix–turn–helix motif found in many DNA-binding proteins (Holland *et al.*, 1987; Brennan & Matthews, 1989). In particular, GerE is related by sequence to the C-terminal DNA-binding 'effector' domains of a large family of bacterial regulatory proteins referred to as the LuxR or FixJ family (Kahn & Ditta, 1991). The structure of a member of this family, the *E. coli* response regulator NarL, has been solved by X-ray crystallography (Baikalov *et al.*, 1996). The C-terminal DNA-binding domain of NarL shares approximately 39% sequence identity with GerE. As such it provides a potential molecular-replacement search model for the GerE structure determination. Molecular-replacement trials using the appropiate atoms of NarL and other helix–turn–helix DNA-binding proteins with sequence similarity to GerE as search models have not been successful. A search for heavy-atom derivatives is now

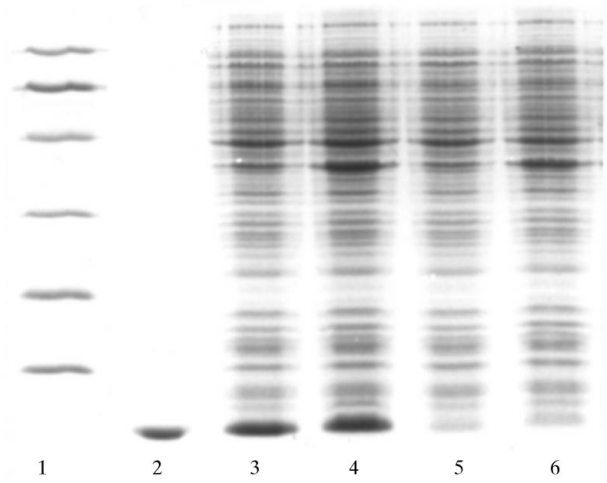

Fig. 1. 17% SDS–PAGE analysis of GerE over-expression. Lane 1, molecular-weight markers (Biorad; 97.4, 66.2, 45, 31, 21.5 and 14.4 kDa). Lane 2, purified GerE. Lanes 3–6, BL21 (DE3) containing GerE expression clone; lane 3, plus IPTG (soluble protein); lane 4 (total protein); lane 5, uninduced sample (soluble protein); lane 6 (total protein).

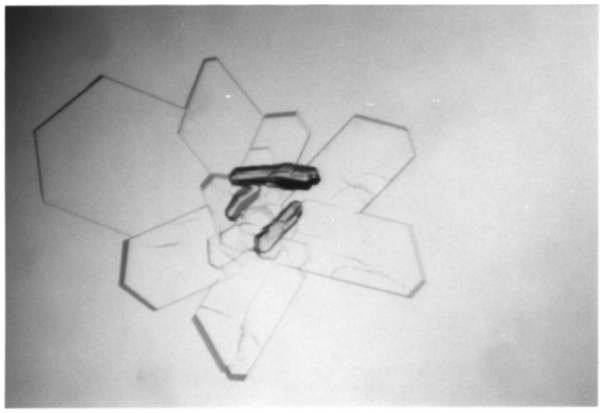

Fig. 2. Crystals of *Bacillus subtilis* GerE. Typical dimensions are 0.6 × 0.2 × 0.05 mm.

underway in order to aid structure determination. The crystal structure of GerE is the first step towards understanding the mechanism of action of this important regulatory system. It will provide a framework for the analysis of DNA binding and help to explain how, through its interaction with σ^K, GerE modulates expression of the coat protein genes during bacterial sporulation.

This work was funded, in part, by the Wellcome Trust (047031/Z/96/Z) and the European Union (through provision of a TMR fellowship to VD) and the Biotechnology and Biological Research Council. The authors would like to thank Gideon Davies and the staff of the CCLRC Daresbury Synchrotron Radiation Source for invaluable assistance.

References

Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R. P. & Dickerson, R. E. (1996). *Biochemistry*, **35**, 11053–11061.

Brennan, R. G. & Matthews, B. W. (1989). *J. Biol. Chem.* **264**, 1903–1906.

Collaborative Computational Project, Number 4. (1994). *Acta Cryst.* D**50**, 760–763.

Cutting, S. & Mandelstam, J. (1986). *J. Gen. Microbiol.* **132**, 3013–3024.

Cutting, S., Panzer, S. & Losick, R. (1989). *J. Mol. Biol.* **207**, 393–404.

Errington, J. (1993). *Microbiol. Rev.* **57**, 1–33.

Feng, P. & Aronson, A. I. (1986). *Curr. Microbiol.* **13**, 221–226.

Henriques, A. O., Beall, B. W. & Moran, C. P. Jr (1997). *J. Bacteriol.* **179**, 1887–1897.

Holland, S. K., Cutting, S. & Mandelstam, J. (1987). *J. Gen. Microbiol.* **133**, 2381–2391.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Jenkinson, H. F. & Lord, H. (1983). *J. Gen. Microbiol.* **129**, 2727–2737.

Kahn, D. & Ditta, G. (1991). *Mol. Microbiol.* **5**, 987–997.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azvedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, C., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, S. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S.-K., Codani, J.-J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhöft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabrer, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.-F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.-M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, M., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.-H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M. Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B.-S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemura, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.-F., Zumstein, E., Yoshikawa, H. & Danchin, A. (1997). *Nature (London)*, **390**, 249–256.

Losick, R. & Stragier, P. (1992). *Nature (London)*, **355**, 601–604.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Moir, A. (1981). *J. Bacteriol.* **146**, 1106–1116.

Otwinowski, Z. (1993). In *Data Collection and Processing: Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. Bailey. Daresbury: Science and Engineering Research Council.

Sacco, M., Ricca, E., Losick, R. & Cutting, S. (1995). *J. Bacteriol.* **177**, 372–377.

Sandman, K., Kroos, L., Cutting, S., Youngman, P. & Losick, R. (1988). *J. Mol. Biol.* **200**, 461–473.

Stragier, P. & Losick, R. (1996). *Annu. Rev. Genet.* **30**, 297–341.

Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.

Zhang, J., Ichikawa, H., Halberg, R., Kroos, L. & Aronson, I. (1994). *J. Mol. Biol.* **240**, 405–415.

Zheng, L., Halberg, R., Roels, S., Ichikawa, H., Kroos, L. & Losick, R. (1992). *J. Mol. Biol.* **226**, 1037–1050.

Zheng, L. & Losick, R. (1990). *J. Mol. Biol.* **212**, 645–660.